Convolutional Neural Network Model for Diabetic Retinopathy Feature Extraction and Classification

Sharan Subramanian[1] and Leilani H. Gilpin[#]

[1]Merrill F. West High School, USA
[#]Advisor

ABSTRACT

The application of Artificial Intelligence in the medical market brings up increasing concerns but aids in more timely diagnosis of silent progressing diseases like Diabetic Retinopathy. In order to diagnose Diabetic Retinopathy (DR), ophthalmologists use color fundus images, or pictures of the back of the retina, to identify small distinct features through a difficult and time-consuming process. Our work creates a novel CNN model and identifies the severity of DR through fundus image input. We classified 4 known DR features, including micro-aneurysms, cotton wools, exudates, and hemorrhages, through convolutional layers and were able to provide an accurate diagnostic without additional user input. The proposed model is more interpretable and robust to overfitting. We present initial results with a sensitivity of 97% and an accuracy of 71%. Our contribution is an interpretable model with similar accuracy to more complex models. With that, our model advances the field of DR detection and proves to be a key step towards AI-focused medical diagnosis.

Introduction

Medical diagnosis serves as the foundational step in patient care, determining the pathway for treatment and intervention. According to Dr. Edmund Arthur of the University of Alabama at Birmingham School of Optometry, the accuracy and timeliness of diagnoses are essential, as they directly influence the patient's prognosis and quality of life[1]. A delay in identifying a condition can cause health issues, increase costs, and lower chances of full recovery. Recognizing a disease in its nascent stages often provides a broader array of treatment options and increases the likelihood of a favorable outcome. Diabetic Retinopathy (DR) exemplifies conditions where early detection can make a profound difference. DR, a complication resulting from diabetes, threatens the retina's blood vessels, jeopardizing the crucial light-sensitive layer at the back of the eye. Despite the potential to prevent up to 98% of vision loss with timely intervention, its silent progression often eludes timely detection, resulting in irreversible vision impairments.

Artificial Intelligence (AI) has emerged as a solution for more effective diagnostic tools. Within AI, there is a subset of complex models called Convolutional Neural Networks (CNNs). However, CNNs applications for DR therapy have been of high complexity, requiring a high amount of computational resources. The most widely used approach for DR diagnosis consists of a dilated eye exam administered by an ophthalmologist or optometrist. However, deep learning models perform with the same accuracy as medical professionals at processing and analyzing fundus images[2], which depict the retina in detail. CNNs may also be able to diagnose other conditions, including

---

[1] Hutton, D., & Harp, M. D. (2023, April 25). Arvo live: Early detection of diabetic retinopathy in underserved communities. Ophthalmology Times. https://www.ophthalmologytimes.com/view/arvo-live-early-detection-of-diabetic-retinopathy-in-underserved-communities

[2] Charters, L. (2023). Finding a place for AI, machine learning in retinal imaging. Www.ophthalmologytimes.com, 48. https://www.ophthalmologytimes.com/view/finding-a-place-for-ai-machine-learning-in-retinal-imaging

cataracts, glaucoma, and even illnesses outside the retina. In Figure 1, there are four main features from a fundus image showing positive signs of DR. By processing these images, CNNs can identify early indicators of DR, illuminating proactive medical interventions aided by AI.
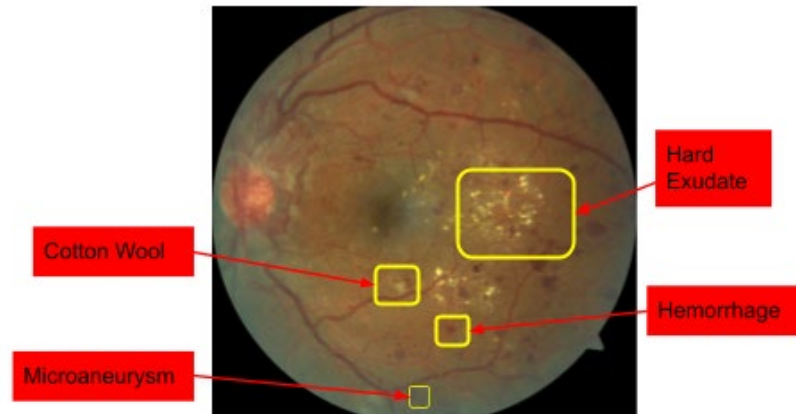
Figure 1. Different features of a CFP show the presence of DR. This Fundus Image was pulled from the Messidor-2 dataset and had a classified severity of 6 (PDR). These are the 4 features that were extracted and processed by our machine-learning model.

In this paper, we present the different models currently available for DR detection and their limitations and challenges. We review the technical methods used, including an in-depth explanation of the CNN architecture, our new methodology and model, and an open-source GitHub repository for public access[3]. The training process and post-training techniques used to improve accuracy are thoroughly examined in the *Explanation of CNN Layers* and *Methods* sections, respectively. In addition to the methodology, we compare pre-trained models with the proposed ADL-CNN model and examine the benefits and disadvantages of each. Furthermore, we explore the potential implications of CNN advancements in healthcare, suggest areas for further research, and discuss the impact these innovations can have on disease detection and management.

Literature Review

We provide a brief history of DR detection and how its diagnosis evolved. In addition to going over the various pre-existing studies on CNN models for DR detection, we also highlight each model's and method's limitations. Then, we will highlight what our model brings that is different from other studies and the advantages our model entails.

Historical Overview of DR Detection

In the early stages of DR diagnosis, its approach was based on a patient's symptomatic presentation and rudimentary examination techniques (Wolfensberger & Hamilton, 2001). Eduard Jaeger and Albert von Graefe, in the 1850s described diabetic macular changes: a historical review suggests the evolution of our understanding (Wolfensberger & Hamilton, 2001). It took more than a century for a consensus to emerge regarding the link between cystoid macular edema, which is caused by the accumulation of fluids in the macula, and diabetes mellitus (DM), where the body's

---

[3] https://github.com/s21sharan/CNN_DR_Detection Github repository with the code for the proposed CNN model. This repository also has a detailed readme file on how to set up and use the model.

immune system attacks and destroys insulin-producing cells in the pancreas. Recognizing the connection between cystoid macular edema and diabetes mellitus today influences diagnostic methods for patients with visual symptoms.

The first line of diagnosis typically involved direct ophthalmoscopy, a method discussed in the historical evolution of DR diagnosis (Matuszewski et al., 2017). Digital fundus photography reshaped DR screening in the late 20th and early 21st century. Comparisons between singly, two-field, and three-field 45-degree Color Fundus Photographs (CFP) showed varying sensitivities and specificities. These advancements have modernized DR diagnosis, with varying sensitivities and specificities observed across different Color Fundus Photograph methods.

With more advances in integrated AI and DR screening, the FDA approved an autonomous AI device that used 45-degree digital CFP for DR detection called the EyeArt AI system. On a side note, confocal scanning ophthalmoscopy emerged as a technique that provided ultra-widefield imaging.

With these advancements, other challenges lie ahead on this path of innovation and integration. The intricacies of integrating various technologies seamlessly, ensuring universal accessibility to these advanced diagnostic tools, and refining AI algorithms for increased accuracy are among the ongoing challenges. While previous models may have offered solutions to specific challenges, a collaborative effort across medical, technological, and regulatory domains is essential for realizing the full potential of these novel diagnostic approaches.

Specific Studies on DR Detection using CNNs

In a study by Yasashvini R. et al., researchers attempted to classify DR using both CNN and hybrid deep CNNs (Yasashvini R. et al., 2022). Their methodology combined the strengths of CNN architectures with a hybrid deep CNN via a mixture of different classifiers. They hypothesized that by combining the power of both these mixed classifiers, the accuracy of DR would increase, better accounting for the variety of features and abnormalities present in DR-affected CFPs. Their results were successful in achieving accuracy rates close to 100%.

Following the onset of high precision in DR-prediction models, Pratt et al. introduced the task of predicting severity in DR diagnosis (Pratt et al., 2016). Their multi-dimensional data supports severity prediction in CFPs, from retinal features such as micro-aneurysms, exudates, and hemorrhages. Pratt and his team developed a CNN architecture that underwent data augmentation to achieve this dual objective. In Figure 2, we can see the architecture that Pratt et al. used. Their CNN model classified the fundus images between the seven different severities with a sensitivity of 95% and an accuracy of 75% on the validation images. In our work, we augment their architecture's weights and layers.
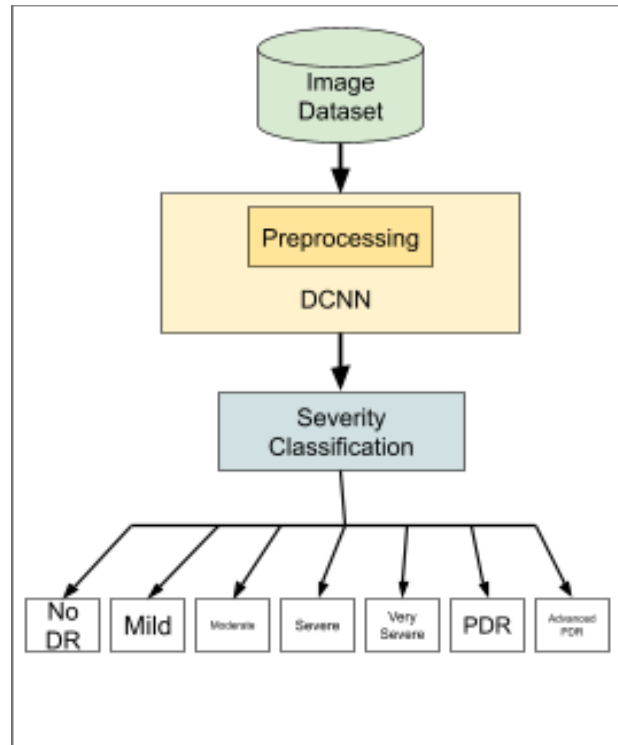
Figure 2. CNN architecture for a DR detection model. This model finds the severity of DR through 7 different classifications of DR (No DR, Mild DR, Moderate DR, Severe DR, Very Severe DR, PDR, and Advanced PDR).

Albahli & Yar's research emphasized the customizability of CNNs in the medical domain (Albahli & Yar, 2021). Albahli & Yar designed three deep-learning models to detect the severity of diabetic retinopathy from retina images and determine its potential progression to macular edema. Their main challenge was the limited dataset they used: the Indian Diabetic Retinopathy Image Dataset (IDRiD). To counteract this, they designed features such as Brightness, Color, and Contrast enhancement, Color Jitters, and Contrast Limited Adaptive Histogram Equalization to generate a broader range of images. Their study used pre-trained models like ResNet50, VGG16, and VGG19 and customized them to DR detection. Their results, after validation, showcased the potential of custom CNNs, with each model yielding reasonable outputs.

On the other hand, the study by Shu-I Pao et al. used entropy images to increase DR detection performance(Shu-I Pao et al., 2020). In this context, the entropy image is an abstract representation computed using the green component of the CFP. By preprocessing these entropy images using unsharp masking, the researchers were able to enhance the clarity and definition of the features. The outcome was a bi-channel CNN, which effectively utilized the features from the gray level and the green component's entropy images. This approach, while complex, provided a comprehensive manner for improving the detection performance of severe DR cases.

In summary, these studies show the immense potential and versatility of CNNs in DR detection. Whether it is the combined powers of CNNs and hybrid networks, the emphasis on severity classification, the adaptability of custom models, or the complex manipulation of entropy images, the domain of DR detection is witnessing fast progress and growth.

Limitations of Current Models

One explicit limitation emerges from the study by Yasashvini et al. While integrating standard CNNs with hybrid deep convolutional networks might enhance precision, it inadvertently adds to the model's complexity. Such sophisticated models often require vast computational resources, both in terms of training and inference. Furthermore, the more complex a model is, the higher the likelihood of overfitting, especially when the available dataset is not large enough. Our proposed model is less complex and has repeatedly been tested on overfitting curves. Overfit models might perform exceedingly well on training data but falter when faced with unseen data, making them unreliable for real-world applications. Testing the model with multiple data sets to ensure uniform reliability in the real world is important.

In the research led by Pratt et al., the results were primarily based on one specific dataset: the Kaggle dataset. The variability and uniqueness of retinal images suggest that a model trained on one dataset might not generalize well to images from different sources, patients of varied demographics, or images captured using different equipment. This bias-related limitation ties into the data scarcity in DR and fundus imaging.

Albahli's research, which introduced custom CNN models, highlighted another significant limitation related to data scarcity. The requirement to employ image generation techniques like BCC enhancing, CJ, and CLAHE to expand their dataset indicates the prevalent challenge of limited datasets in the field. Having a limited dataset not only hampers the depth of training but also raises concerns about the model's ability to generalize to a broader range of images. In our model, we have trained and tested it with a more comprehensive dataset and have seen repeatable and reliable results with additional data sets. Moreover, artificially augmenting data can sometimes introduce new artifacts or fail to replicate the subtle nuances of natural images, potentially leading to models that might misinterpret or overlook certain features. Another question arises by introducing many data augmentation techniques: whether pre-processing techniques mask significant data features.

Shu-I Pao and his team's research brings forth the intricacies of preprocessing. While the emphasis on entropy images derived from the green component of fundus photographs and preprocessing techniques like unsharp masking is innovative, it also adds layers of processing that might not be feasible in real-world, time-sensitive scenarios. Such intricate preprocessing might also introduce biases or inadvertently filter out certain relevant features, limiting the model's diagnostic capabilities. Our model sticks to basic preprocessing to resize the image and extract key features without modifying or filtering the actual image itself.

Lastly, a collective limitation observed from all these studies is the gap between theoretical performance and real-world implementation. While these models showcase impressive metrics in controlled environments, challenges such as computational constraints, integration with existing healthcare systems, interpretability of results, and ensuring consistent performance across diverse patient populations remain. Addressing these practical limitations is paramount for the successful and widespread adoption of CNNs in DR diagnostics.

AI Trust in Healthcare Environment

The area of healthcare, particularly in recent years, has observed a large fusion of artificial intelligence (AI) applications, covering a broad spectrum ranging from diagnostics and clinical decision-making to digital public health. AI's potential to enhance clinical outcomes and efficiency is acknowledged, but the acceptance of this technology remains contingent on multiple factors.

During the COVID-19 pandemic, Chalutz Ben-Gal's study investigated the acceptance of AI in primary care (Chalutz Ben-Gal, 2022). It was found that while many patients showcase a resistance towards AI, several factors can potentially tilt the balance in favor of AI. Explicit mentions of AI's superior accuracy, positive nudges from primary care physicians, and assurances of personalized patient listening experiences could enhance AI acceptance. Notably, Robertson et al. found that specific demographics, including Black respondents and older individuals, were less

inclined to choose AI, highlighting the consideration of demographic-specific concerns in AI adoption (Robertson et al., 2023).

Diving deeper into the diagnostic field, the potential of AI to augment diagnostic precision is immense. However, as noted by Robertson et al., patients' trust in AI may waver based on disease severity and their perceived personalization of the AI system.

Lastly, considering the viewpoint of healthcare professionals, Lambert et al. presented an integrative review highlighting the barriers to AI acceptance in the hospital setting (Lambert et al., 2023). The fear of losing professional autonomy in integrating AI into existing clinical workflows was a recurring concern. Conversely, training geared towards AI utilization emerged as a positive catalyst and gained acceptance among countless professionals. The study further emphasized the significance of involving end-users during the early stages of AI development, suggesting a collaborative approach to ensure that the technology meets real-world needs and challenges.

In conclusion, while the potential benefits of AI in healthcare are vast, its acceptance remains an issue, necessitating a comprehensive understanding of both patients' and practitioners' perspectives coupled with tailored approaches to address their specific concerns.

Understanding Convolutional Neural Networks

In-Depth Explanation of CNNs

Convolutional Neural Networks (CNNs) are pivotal in machine learning and artificial intelligence, especially for image processing and classification tasks. Inspired by biological neural networks, CNNs emulate human perception by recognizing patterns from image pixels, enabling classification, identification, and visual data reconstruction (Shin et al., 2016). In the realm of healthcare, particularly radiology, they exhibit considerable potential in enhancing conventional diagnostic methodologies, resulting in improved speed, reliability, and occasionally superior accuracy (Yamashita et al., 2018).

Fundamentally, neural networks comprise interconnected nodes or "neurons" that execute elementary mathematical operations. These networks are constructed to model intricate data relationships. As depicted in Figure 3, CNNs encompass input layers for data reception, hidden layers for data processing, and output layers for prediction/decision-making. Unlike conventional neural networks employing full interconnections between layers, CNNs employ sparse connections to focus on local features, rendering them highly efficient in capturing spatial hierarchies. Consequently, CNNs are exceptionally well-suited for tasks such as image recognition, where the extraction of local features such as edges, corners, and textures offers valuable information.
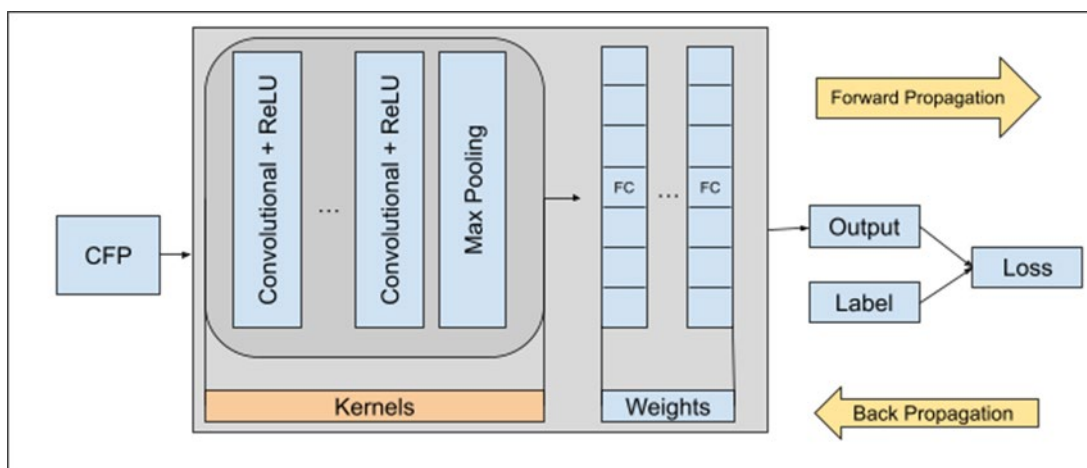
Figure 3. This flowchart highlights the CNN architecture that is used in our proposed model. It has different kernels and depth layers that take input images and output the classification. The input layer is where the color fundus image is fed into the neural network. Then the convolutional layers apply a set of filters to the input image, detecting specific features like edges, textures, and shapes that are tied into the DR signs, as highlighted in *Figure 1*. Then, the ReLU activation function is applied element-wise, introducing nonlinearity. The weights represent the parameters that the network learns during training, and the final output is the severity classification of DR based on features extracted from the input image by preceding layers.

CNNs belong to a category of neural networks known as Deep Learning models (Alzubaidi et al., 2021). As the name suggests, these models are characterized by their depth or the number of layers data must pass hierarchically. Deep Learning, a subfield of machine learning, offers significant advantages over traditional machine learning methods. The depth of the network allows for more complex features to be learned, thereby increasing the model's ability to understand intricate patterns in the data, especially in medical imaging tasks (Yala et al., 2019).

Breaking Down the Architecture: Layers, Filters, Feature Extraction, etc.

The convolutional layer is the core aspect of a CNN. This layer performs the convolution operation on the input, passing the result to the next layer. These operations focus on local regions of the input, preserving the relationship between pixels by learning image features (Taye et al., 2023). For instance, in DR detection, this localized focus enables CNNs to discern pixel-level details, capturing critical retinal features and enhancing the model's feature-learning capability.

After convolution, the layers often employ an activation function, the most common being the Rectified Linear Unit (ReLU) (Indolia et al., 2018). ReLU, in the kernels section of *Figure 3,* introduces non-linearity into the system, which allows the model to learn from the error and make adjustments, improving its performance during the training process.

Following the convolutional and activation layers, CNNs usually contain pooling layers to reduce dimensionality. The pooling operation condenses the feature map obtained from the previous layer, usually by taking the maximum (as in *Equation 1*) or average value (as in *Equation 2*) from a group of values in the feature map[4]. This pooling process also results in reduced size and computational complexity of the resultant feature map.

Equation 1: Maximum Pooling

$$Max\ Pooling(x,y)\ =\ argmax_{i,y}(Input(x+i,y+j))$$

Equation 2: Average Pooling

$$Average\ Pooling(x,y)\ =\ \frac{1}{n^2}\sum_{i=0}^{n-1}\sum_{j=0}^{n-1}(Input(x+i,y+j))$$

One of the significant operations in CNNs involves filter or kernel feature extraction, as illustrated in Figure 3. Filters, which are small and adjustable parameters, traverse the input image, generating feature maps. For example, one filter may focus on edge detection, while another specializes in identifying specific colors. The concept of "stride" defines how filters move across the input image, determining their step size. These adjustable parameters collectively provide spatial control over the convolution operation, enabling precise adjustments and customization.

---

[4] Nanos, G., & Nanos, G. (2023). Neural Networks: Pooling layers | Baeldung on Computer Science. Baeldung on Computer Science. https://www.baeldung.com/cs/neural-networks-pooling-layers

Towards the end of the network, fully connected layers, also known as dense layers, utilize the high-level features acquired from preceding layers to classify the image into distinct categories. These layers share a structural resemblance to a standard multi-layer perceptron neural network.

Explanation of CNN Layers, with a Specific Focus on Image Classification Tasks

Forward propagation in CNNs entails the sequential transmission of the input image through multiple layers. As the image traverses from one layer to the next, it gives rise to feature maps characterized by reduced dimensionality, which are subsequently employed in the final layers for classification purposes (Peng et al., 2021). Similar to other neural networks, CNNs use backpropagation for training. This algorithm evaluates the disparity between the output produced by forward propagation and the expected outcome, subsequently modifying the filter weights, as illustrated in Figure 3. This weight adjustment process is often facilitated by Stochastic Gradient Descent (SGD) optimizers.

1. Initialization: Initialize the weights (w_1, w_2) and biases (b_1, b_2) with random values.
2. Forward Pass: Pass a training example (x_{train}) through the network to compute the predicted output (y)
3. Compute Loss: Calculate the loss, as in equation 3, using predicted output and the true target

   Equation 3: Loss function using Mean Squared Error (MSE)

$$L = \frac{1}{2}(y - y_{true})^2$$

4. Backpropagation: Compute the gradients of the loss with respect to the weights and biases using the chain rule (calculus)
5. Update Weight: Adjust the weights using the gradients and small learning rate according to SGD update rule[5], as in Equation 4 and 5

   Equation 4: Weight Update Equation

$$w_i = w_i - \alpha\left(\frac{\vartheta L}{\vartheta w_i}\right)$$

   Equation 5: Bias Update Equation

$$b_i = b_i - \alpha\left(\frac{\vartheta L}{\vartheta b_i}\right)$$

6. Repeat: Repeat steps 2-5 for specified number of epochs until convergence

Pre-trained models are applied in various scenarios, particularly when working with limited training datasets. This approach applies pre-trained CNNs, with subsequent fine-tuning of their final layers for adaptation to new tasks. This reduces computational resource demands instead of using significant compute to train a CNN from scratch. CNNs utilize batch processing techniques, which enhance training efficiency by enabling faster and more parallel computation. Furthermore, cross-entropy loss functions are used to quantify the disparity between predicted outputs and actual labels, with the primary training objective being the minimization of this loss function. Without the cross-entropy loss function, CNNs risk overfitting, introducing biases, and failing to generalize effectively on unseen data. Similarly, a CNN's performance can be influenced by hyperparameters like learning rate, filter size, filter count, and layer count. Consequently, hyperparameter optimization is frequently conducted to refine these settings.

---

[5] Ruder, S. (2020). An overview of gradient descent optimization algorithms. ruder.io. https://www.ruder.io/optimizing-gradient-descent/

Beyond image classification, CNNs find applications in complex tasks such as video analysis, natural language processing, and autonomous driving. In the healthcare domain, their image recognition capabilities are invaluable for the detection of anomalies in X-rays, MRIs, CT scans, and CFPs, as elaborated in this paper. As CNNs and their associated technologies continue to evolve, they hold immense promise for healthcare applications, ranging from early disease classification for conditions like cataracts and cancers to real-time patient monitoring.

Methodology

Data Sets Used

Table 1. The different publicly available CFP datasets for classifying DR.

Dataset	Year Released	Image Count	Field of Vision (FOV)
Kaggle	2015	88k	50°
APTOS	2019	13k	Not specified
DDR	2019	13.6k	45°
Messidor-2	2010	1748	45°
DeepDRid	2019	2256	Not specified

We incorporated multiple data sets to develop, train, and validate our model, with a primary focus on the Messidor-2 dataset. Table 1 gives a snapshot of the datasets we utilized.

The primary dataset which our model was built from is the Messidor-2 dataset. Between October 2009 and September 2010, diabetic patients were imaged using a Topcon TRC NW6 non-mydriatic fundus camera with a 45° field of view, resulting in 345 total DR examinations. Only macula-centered images were considered for this dataset. This was then combined with the Messidor-original database, which has a total of 529 DR examinations, to create the Messidor-2 Data set.

Initially, we used a 70-30 train-test split with the Messidor-2 dataset taking inspiration from other studies in the field (Mutawa et al., 2023). However, we adjusted this split to 80-20 to minimize training and validation losses. Our secondary data sets include the Kaggle dataset, APTOS, DDR, and DeepDRid.

1. Kaggle: This dataset was involved in one of the Kaggle Challenges relating to medical imaging and CNNs. This dataset covers a diverse range of sources and conditions, with over 88,000 images.
2. APTOS (Asia Pacific Tele-Opthalmology Society): This society concentrates on eye-related conditions, including DR. This dataset includes patients of Asian descent and is very comprehensive, with over 13k images.
3. DDR: This dataset is sourced from 147 hospitals spread across 243 provinces in China. These images are classified based on the severity of diabetic retinopathy into 5 categories.
4. DeepDRid: This dataset originated from the 2nd Diabetic Retinopathy: Segmentation and Grading Challenge in collaboration with ISBI in 2018. This dataset focuses on three pivotal tasks: dual-view disease grading, image quality estimation, and transfer learning.

All these datasets were classified based on the severity of DR present in the images. The five severities were none, mild, moderate, severe, and proliferative DR. The images of poor data quality were excluded from the datasets, and a manual review took place before putting these images into our model. In Figure 4, we can see examples of CFPs

that passed this review and made it onto the dataset. There are significant differences between these CFPs and their features, including the hard exudate, cotton wool, and hemorrhage, as annotated in Figure 1.
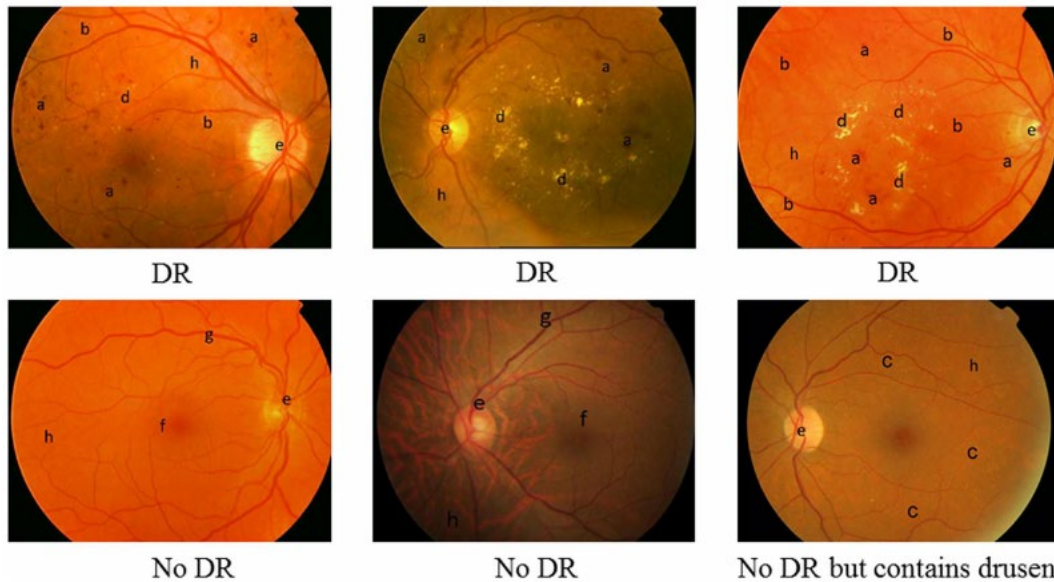
Figure 4. This figure has three images of CFPs that contain DR (in the top row) and three images of CFPs that do not contain DR (in the bottom row). The annotated features are as follows: a. Hemorrhage; b. MA; c. Drusen; d. Exudates; e. Optic disc; f. Fovea; g. Blood vessel; h. Background doi: https://doi.org/10.1371/journal.pone.0066730.g001

Methods

Our approach is split into two primary stages: the training/validation phase and the testing phase.

Data was aggregated from various sources like hospitals and specialized data repositories during the initial phase. We used the Messidor-2 dataset as our baseline data. We removed any 25 low-quality images from the consolidated dataset through an ancillary intelligent model and organized them using a 70-30 split[6]. We split each subcategory, or severity, using Sklearn's train-test split function.

During the pre-processing stage, we used two operations to reduce noise and remove gaps in the background and foreground of the CFP. The first procedure is known as Erosion. Erosion is used to eliminate or spike the edge of the area and is represented by Equation 1(Sarki et al., 2021).

Equation 6: Erosion Equation

$$A \ominus B = \{p | B_p \subseteq A\}$$

Then we also applied dilation, which is used to broaden the rim of the background or foreground image configuration. This procedure helps us fill the gap created through erosion and is defined by this equation (Sarki et al., 2021):

Equation 7: Dilation Equation

$$A \oplus B = \{x | B_x \cap X \neq 0\}$$

---

[6] Our updated dataset is attached to the GitHub link in *Footnote 3*.

Where $\oplus$ denote the dilation; $\ominus$ denote the erosion; A = Structuring element and B = the erosion of the dilation of that set. These equations are from the Tyler Coye Algorithm.

In the CNN model, we used convolutional operations to extract features, as highlighted in Figure 5. We could convert the images into feature activation maps using diverse filter sizes. Then, using the Rectified Linear Unit (ReLU) activation function, we further enhanced this model with non-linearity. Following this, we employed a pooling layer, an instrumental part in compressing the activation map's dimensions without significant information loss. We then vectored the activation maps, processed through several fully connected layers, resulting in a severity rating based on the DR-positive or DR-negative classification.
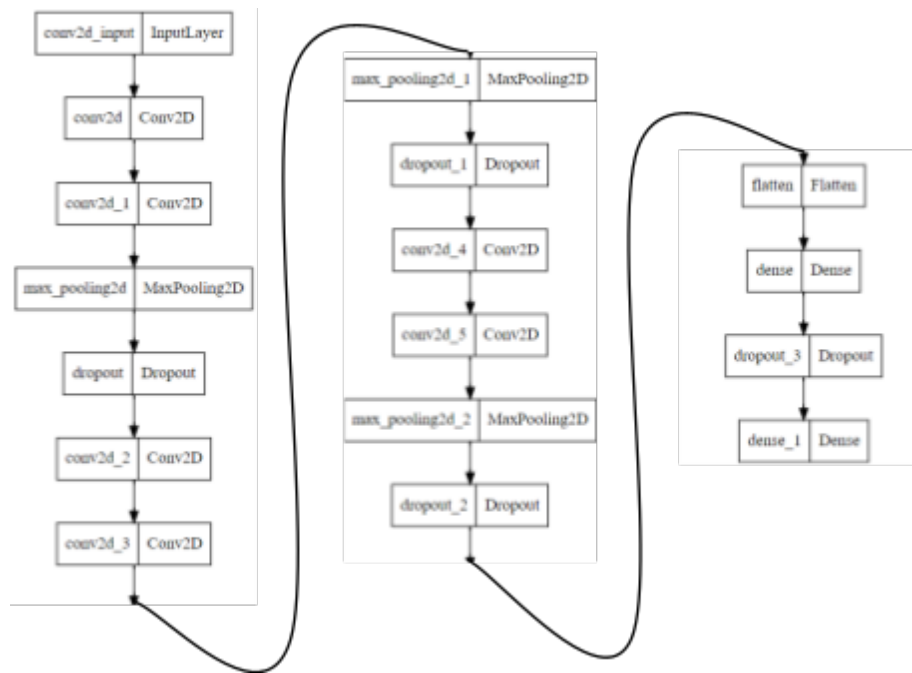
Figure 5. This flowchart represents the different layers of our model as depicted in a sequential by Keras. There are a total of 6,566,055 trainable parameters in our model, with each spread across the layers in the figure.

In the model's second stage, we tested the data using our test-split of the Messidor-2 data set and other secondary data sets, as highlighted in section (I). To prevent our model from overfitting, we implemented the early stopping mechanism, which was used to reduce training time from 8 hours 36 mins to 4 hours and 55 mins. This technique monitors our model's performance on the validation set and halts training once the performance starts deteriorating. This ensures that the model does not overfit. We also implemented a drop-out mechanism, which allowed randomly selected neurons to be ignored at each iteration. This ensured that our neural network remained accurate and generalized well to new-unseen data.

We also used the Keras auto-hyperparameter tuning module to tune crucial parameters of our models. This algorithm uses a mathematical approach to parameter tuning; this tweaks the learning rate, batch size, epochs, and dropout rate to ensure the lowest validation losses and highest accuracy. We were able to increase our validation accuracy from 55% to 70% using this hyperparameter tuner.

Results

In our initial study, the DR model was trained, validated, and tested on the Messidor data set. With 1748 images split into five severity categories, we had about 300 for each severity. These 500 images were then split in an 80-20 ratio for the test-train split.

As seen in Figure 6a, our model's performance throughout the training process is consistent. Averaging at approximately 97% over 100 epochs, this model can recognize and predict patterns within the training dataset. The validation accuracy remained at around 70% throughout the 100 epochs. Though slightly lower than the training data, this validation accuracy aligns with the understanding that generalizing our model to unseen data will reasonably lower the accuracy.
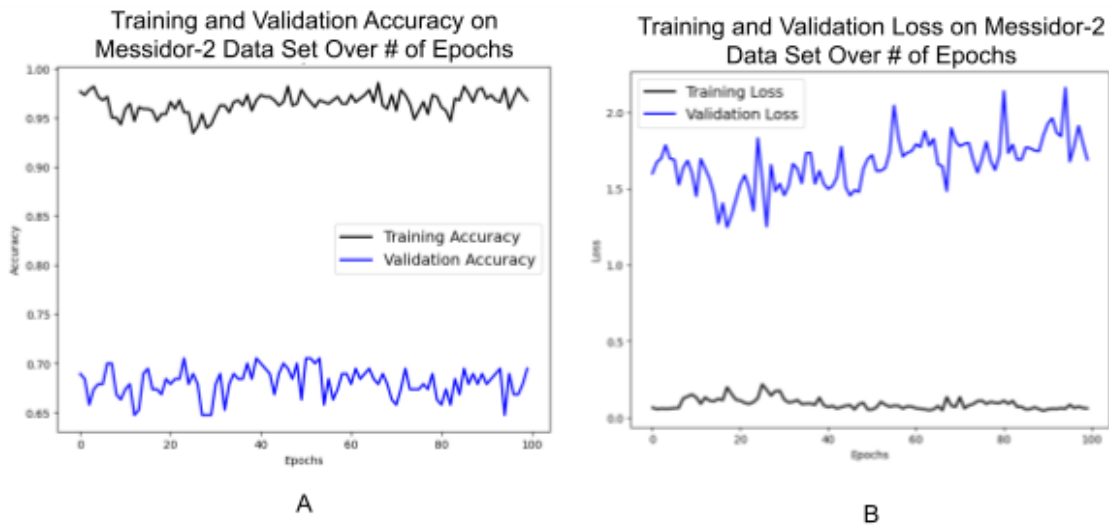
Figure 6. Figure 6a shows the Accuracy curves for our model as tested on the Messidor-2 Data set. Figure 6b shows the Loss curves for our model as tested on the Messidor-2 Data set.

The training loss, as seen in Figure 6b, remained under 0.3 across all epochs. On the other hand, the validation loss was higher at about 1.5-1.8. As shown in Table 2, the training and validation accuracy also stabilized around the seventh epoch, hinting that our accuracies were not just a fluke. Additionally, this model was reset and repeatedly tested, bringing stable and similar accuracy values every time.

Table 2. Training and Validation Results for Proposed CNN model

Epochs	Training Loss	Training Accuracy	Validation Loss	Validation Accuracy
0	0.075937	0.971781	1.577781	0.700000
1	0.076237	0.964727	1.689509	0.694737
2	0.083331	0.970018	1.463975	0.710526
3	0.076263	0.973545	1.697605	0.700000

Epochs	Training Loss	Training Accuracy	Validation Loss	Validation Accuracy
4	0.96008	0.962963	1.778189	0.663158
5	0.073751	0.966490	1.582166	0.693684
6	0.072941	0.964727	1.623718	0.784211
7	0.076225	0.966490	1.609345	0.700000

Discussion

Our model demonstrates the effectiveness of CNNs in DR detection, with a streamlined architecture. In comparison to Yasashvini R. et al.'s hybrid deep CNNs, our model strikes a balance between complexity and efficacy. Unlike Albahli et al.'s research, which relied on multiple image enhancement techniques due to data scarcity, our model was trained on a more comprehensive dataset and validated on additional datasets.

In contrast to techniques like entropy-based image processing highlighted by Shu-I Pao et al., our model aims for direct, effective, and reproducible DR classification, maintaining competitive validation accuracy.

A 20% delta between training and validation metrics indicates room for improvement. Future work includes implementing regularization techniques like L1 or L2 regularization to address this disparity. Additionally, inspired by Su-I Pao et al.'s use of entropy images, we may explore further feature engineering techniques to extract informative features from retinal images.

In the evolving field of DR detection, our model offers promise and scalability. Future iterations, along with advancements such as regularization techniques, have the potential to improve accuracy. Our approach's simplicity enables integration with diverse datasets and applicability across demographics and equipment sources.

Conclusion

In our study on DR detection using CNNs, we achieved enhanced efficiency and accuracy in categorizing DR severity levels from retinal images. Our approach, which incorporated preprocessing techniques like erosion and dilation alongside convolutional operations, consistently yielded a 97% training accuracy over 100 epochs. Impressively, our validation accuracy reached 71%, marking significant progress towards automated medical diagnosis in DR.

Comparative analysis with other studies underscores the effectiveness of our approach, striking a balance between model understanding and detection efficacy. The model is adaptable across diverse datasets, requiring minimal image enhancement due to the comprehensive Messidor-2 dataset used. However, the gap between training and validation metrics indicates potential for model enhancement, including advanced regularization techniques and feature engineering to improve accuracy and mitigate overfitting.

Future improvements may explore more advanced preprocessing techniques and incorporate regularization methods to bridge the gap between training and validation loss metrics. Collaborating with healthcare institutions to obtain diverse datasets could further enhance the model's adaptability across various clinical scenarios.

Our consistent results and the interpretability of our approach underscore the reliability of CNNs in DR detection. Our work raises questions about how healthcare professionals will adapt to AI technologies in the medical field and the potential for more optimal Neural Networks in medical classification tasks. As the medical community embraces the intersection of healthcare and AI, the advancements presented in this and other studies contribute to a safer and more clinically advanced society.

Acknowledgments

References

Albahli, S., & Nabi Ahmad Hassan Yar, G. (2022). Detection of Diabetic Retinopathy Using Custom CNN to Segment the Lesions. *Intelligent Automation & Soft Computing*, *33*(2), 837–853. https://doi.org/10.32604/iasc.2022.024427

Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M. A., Al-Amidie, M., & Farhan, L. (2021). Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data*, *8*(1). https://doi.org/10.1186/s40537-021-00444-8

Chalutz Ben-Gal, H. (2023). Artificial intelligence (AI) acceptance in primary care during the coronavirus pandemic: What is the role of patients' gender, age and health awareness? A two-phase pilot study. *Frontiers in Public Health*, *10*. https://doi.org/10.3389/fpubh.2022.931225

Indolia, Sakshi & Goswami, Anil & Mishra, S.P. & Asopa, Pooja. (2018). Conceptual Understanding of Convolutional Neural Network- A Deep Learning Approach. Procedia Computer Science. 132. 679-688. https://doi.org10.1016/j.procs.2018.05.069

Matuszewski, D. J., Hast, A., Wählby, C., & Ida-Maria Sintorn. (2017). A short feature vector for image matching: The Log-Polar Magnitude feature descriptor. *PLOS ONE*, *12*(11), e0188496–e0188496. https://doi.org/10.1371/journal.pone.0188496

Mesfin Asfaw Taye, Morrow, D., Cull, J., D. Hudson Smith, & Hagan, M. T. (2022). Deep Learning for FAST Quality Assessment. *Journal of Ultrasound in Medicine*, *42*(1), 71–79. https://doi.org/10.1002/jum.16045

Mutawa, A. M., Alnajdi, S., & Sruthi, S. (2023). Transfer Learning for Diabetic Retinopathy Detection: A study of dataset combination and model performance. Applied Sciences, 13(9), 5685. https://doi.org/10.3390/app13095685

Pao, Shu-I., Lin, H.-Z., Chien, K.-H., Tai, M.-C., Chen, J.-T., & Lin, G.-M. (2020). Detection of Diabetic Retinopathy Using Bichannel Convolutional Neural Network. *Journal of Ophthalmology*, *2020*, e9139713. https://doi.org/10.1155/2020/9139713

Peng, J., Pedersoli, M., & Desrosiers, C. (2021). Boosting Semi-supervised Image Segmentation with Global and Local Mutual Information Regularization. *Machine Learning for Biomedical Imaging*, *1*(MIDL 2020 special issue), 1–29. https://doi.org/10.59275/j.melba.2021-g79f

Pratt, H., Coenen, F., Broadbent, D. M., Harding, S. P., & Zheng, Y. (2016). Convolutional Neural Networks for Diabetic Retinopathy. *Procedia Computer Science*, *90*, 200–205. https://doi.org/10.1016/j.procs.2016.07.014

R., Y., Raja Sarobin M., V., Panjanathan, R., S., G. J., & L., J. A. (2022). Diabetic Retinopathy Classification Using CNN and Hybrid Deep Convolutional Neural Networks. *Symmetry*, *14*(9), 1932. https://doi.org/10.3390/sym14091932

Robertson, C. T., Woods, A., Bergstrand, K., Findley, J., Balser, C., & Slepian, M. J. (2023). Diverse patients' attitudes towards Artificial Intelligence (AI) in diagnosis. *PLOS Digital Health*, *2*(5), e0000237–e0000237. https://doi.org/10.1371/journal.pdig.0000237

Sarki, R., Ahmed, K., Wang, H., Zhang, Y., Ma, J., & Wang, K. (2021). Image Preprocessing in Classification and Identification of Diabetic Eye Diseases. *Data Science and Engineering*, *6*(4), 455–471. https://doi.org/10.1007/s41019-021-00167-z

Shin, H.-C., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D., & Summers, R. M. (2016). Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning. *IEEE Transactions on Medical Imaging*, *35*(5), 1285–1298. https://doi.org/10.1109/tmi.2016.2528162

Sophie Isabelle Lambert, Madi, M., Sasa Sopka, Lenes, A., Stange, H., Claus-Peter Buszello, & Stephan, A. (2023). An integrative review on the acceptance of artificial intelligence among healthcare professionals in hospitals. *Npj Digit. Med*, *6*(1). https://doi.org/10.1038/s41746-023-00852-5

Wolfensberger, T. J., & Hamilton, A. M. P. (2001). Diabetic retinopathy – An historical review. *Seminars in Ophthalmology*, *16*(1), 2–7. https://doi.org/10.1076/soph.16.1.2.4220

Yala, A., Lehman, C., Schuster, T., Portnoi, T., & Barzilay, R. (2019). A deep learning mammography-based model for improved breast cancer risk prediction. Radiology, 292(1), 60–66. https://doi.org/10.1148/radiol.2019182716

Yamashita, R., Nishio, M., Do, R. K. G., & Togashi, K. (2018). Convolutional neural networks: an overview and application in radiology. *Insights into Imaging*, *9*(4), 611–629. https://doi.org/10.1007/s13244-018-0639-9